An Examination of the Relationship Between Subsidies on Production and Technical Efficiency in Agriculture: The Case of Cotton Producers in Greece.

Grigorios Emvalomatis [a,b], Alfons Oude Lansink [b], and Spiro E. Stefanou [a,b]

[a] **The Pennsylvania State University and** [b] **Wageningen University**

Contact: grigorios.emvalomatis@wur.nl

Paper prepared for presentation at the 107th EAAE Seminar *"Modelling of Agricultural and Rural Development Policies"*. Sevilla, Spain, January 29th -February 1st, 2008

Abstract

Subsidies on production have long been criticized for protecting producers from competition, and thus removing an incentive for efficient use of the resources. This study undertakes the examination of the impact of partially decoupled subsidies on the technical efficiency scores of cotton producers in Greece. The results indicate that compensatory area payments reduce the efficiency scores of the producers by diverting resources from products for which the subsidy is based on the area planted to the production of cotton, for which the aid is related to the volume of output.

Key words: panel data, stochastic distance function, subsidies in agriculture

1. Introduction

The Common Market Organization for cotton was established in 1981, when Greece entered the European Community (EC). It consisted of a guide (target) and a minimal price per tonne of unginned cotton, accompanied by a co-responsibility levy when the production exceeded the Maximum Quantity Guaranteed (MQG) . Originally the MQG was set at 560,000 tonnes and the minimum price at the 95% level of the guide price. The calculation of the support per tonne was made based on the world price of ginned cotton. The difference between the world and the minimal price was paid to the cotton ginners, and from them transferred to the farmers. The same cotton regime applied with minor changes to the other two cotton producing states, Spain and Portugal, when they joined the EC. The major difference was the increase of the MQG to 752,000 tonnes.

During the period 1981-1986 cotton production almost doubled in Greece but always remained below the MQG. After 1986 and until 1995 the production of cotton and the land devoted to cotton production grew at an annual rate of about 10%. At the same time no increasing trends are apparent for Spain, neither in terms of land occupied by cotton nor in terms of production. Due to the increase in production from Greece, the MQG was exceeded every year after 1986.

A slight revision of the Common Market Organization for cotton took place in 1992, but the system underwent its first major revision in 1995. The MQG was increased to 1,031,000 tonnes and divided among the two major producer-states (782,000 tons for Greece and 249,00 tons for Spain). The minimal price was to be reduced only for the country that exceeds its National Quantity Guaranteed (NQG). The minimal price was set at €1,009.9/tonne for the next 5 years and the minimum total amount of aid at the EU level was set at €770 million. Furthermore, quality standards were set for the cotton that is eligible for support, in an attempt to improve the quality that declined since the introduction of the cotton regime.

Even after the 1995 reform, the NQGs were exceeded in both Greece and Spain every year until 2000. Even after the reductions in the minimal price through the co-responsibility mechanism, cotton remained the most profitable crop for farms specializing in the production of arable crops. Karagiannis and Pantzios (2002) present in a simulation study for the case of Greece that since the NQG was not further divided among producers, the individual farmer had an incentive not to comply with the NQG, while compliance would result to a Pareto improvement both for the producers and the taxpayers.

Like most trade distorting measures, the support policies for cotton resulted in a large increase of the produced output and distorted the output mix of the farms. As the price paid to the cotton producers within the EU was well above twice the world price, the price support programs induced them to produce as much output as possible, with little attention paid to the quality of the product. The incentives given to the farmers were further complicated by the 1992 CAP reform, when compensatory area payments were given to the producers of cereals, oilseeds and protein crops, as these arable crops participate in the rotation of cotton.

In 2003 a further revision of the Common Market Organization for cotton was proposed. Under this new regime, a decoupling of aid from production is attempted, with 65% percent of the total aid coming in the form of a single farm payment and the remaining 35% in the form of a payment coupled to cotton area.

The effect of decoupling support from the level of output in agriculture has been studied extensively in the agricultural economics literature. For example, Oude Lansink and Peerlings (1996) and Guyomard et al. (1996) use multiple-output profit functions to study the effect of the compensatory area payments on output levels of arable crops, in the Netherlands and France respectively. Both papers conclude that the compensatory area payments are only partially decoupled. Oude Lansink and Peerlings further conclude that output of arable crops which are not included in the area payments scheme is most likely to increase. Moro, and Sckokai. (1999) following a similar approach, find that the decoupled payments affect output levels mainly through land allocation. Some of the literature focuses on the effect of the policy reforms on output through the reduction of the risk associated with production (Hennessy (1998), Sckokai and Moro (2006)).

Although most studies assessing the impacts of decoupled payments address the impact of land allocation and optimal yield responses, little attention has addressed the impact of policy changes on the efficiency of producers. The impact of subsidies on efficiency has been studied before in the literature (see for example Bezlepkina et al. (2005) and Kleinhanß et al. (2007)). In addition to measuring the presence of inefficiency in production, there is interest in identifying the forces that differentiate efficient from less efficient producers. Stefanou and Saxena (1988), Riefenschnieder and Stevenson (1991), Kumbhakar et al (1991) and Battese and Coelli (1995) are early contributors to literature on explaining efficiency levels. Clearly, managerial ability and related forces can have an impact on the character of efficiency but we are often depending on fairly static measures such as years of education and years of operating experience. A constellation of other forces that are largely unobserved are part of the managerial ability impact on efficiency. The classic fixed effects models were initial efforts to account for unobserved heterogeneity as we estimate the structure of production decision making. More recently, Greene (2005) discusses the random effects models and their variants have been extended to provide greater insight in efficiency measurement to account for the impact of unobserved heterogeneity.

This paper addresses the impact compensatory area payments had on the efficiency of cotton producers in Greece, when the cotton regime consisted primarily of the price support system and the co-responsibility levy. We hypothesize that support schemes that are partially or fully decoupled from production weaken the subsidy related incentives for producers to operate efficiently. Furthermore, when price support and (partially) decoupled payment measures are combined, we would expect the producers to divert non-land resources from the production of commodities for which support is (partially) decoupled from output to those for which it is fully related to output. This would result in distortion of both input and output mix and eventually to a reduction in efficiency. The econometric section of this paper develops a simulation based method to account for the effect of unobserved heterogeneity.

The paper is organized as follows. The next section describes the specification of the model and the econometric technique to account for heterogeneity and to measure the effect of subsidies on technical efficiency when multiple outputs are produced. The data are described in section 3. The results are presented and discussed in section 4. Section 5 concludes the paper.

2. Methodology

Among the numerous techniques proposed in the past for measuring efficiency, the stochastic frontier approach (Aigner et al., 1977; Battese and Corra, 1977; Meeusen and van den Broeck, 1977) appears to be the most popular parametric method. It makes use of a definition of efficiency that uses observed data right from the beginning of the formulation of the estimation problem. It furthermore provides a direct way of determining the effects of exogenous factors on inefficiency.

In a non-stochastic environment and when multiple outputs are produced by a firm, the technical efficiency of firm i in period t is defined as the maximum possible expansion of the observed output vector ($\mathbf{y}_{it}$) such that it reaches the boundary of the production possibilities set, given the inputs used ($\mathbf{x}_{it}$):

$$\mathrm{TE}_{it} = \min\{\mu : \mathbf{y}_{it} / \mu \in P(\mathbf{x}_{it})\} = D_o(\mathbf{x}_{it}, \mathbf{y}_{it}; \boldsymbol{\beta}) \tag{1}$$

Here, $D_o(\mathbf{x}_{it}, \mathbf{y}_{it}; \boldsymbol{\beta})$ represents the deterministic part of an output-oriented distance function and it is assumed to be known up to a vector of parameters $\boldsymbol{\beta}$. This measure of efficiency is bounded between zero and one. In the literature technical efficiency is replaced by $e^{-u_{it}}$, with $u_{it} \geq 0$.

It is apparent from equation (1) that distance functions are linearly homogeneous in outputs. Using this property the last equation can be written as:

$$-\log y_{it}^{M} = \log D_o\left(\mathbf{x}_{it}, \frac{\mathbf{y}_{it}}{y_{it}^{M}}; \boldsymbol{\beta}\right) + u_{it} + v_{it} \tag{2}$$

where y_{it}^{M} is the normalizing output (see Coelli and Perelman (1996) for details). In the last equation v_{it} represents a realization of the random conditions that affect the production process and which are outside the control of the firm.

Once distributional assumptions are placed on v_{it} and u_{it} the parameters in model (2) can be estimated by maximum likelihood. Typically, v_{it} is assumed to be distributed normally, given that it represents random noise in the production process. The distributional assumption on u_{it} is more arbitrary. In the literature, u_{it} has been assumed to have a half-normal (Aigner et al., 1977), truncated normal (Stevenson, 1980), exponential or Gamma distribution (Greene, 1990).

Battese and Coelli (1995), building on the work of Reifschneider and Stevenson (1991) and Kumbhakar et al. (1991), provide an extension where the mean of the distribution of u_{it} depends on a set of firm-specific characteristics ($\mathbf{z}_{it}$). In this model, u_{it} is assumed to have a normal distribution truncated from below at zero, the mean of which is $\mathbf{z}_{it}'\boldsymbol{\xi}$. The vector of parameters $\boldsymbol{\xi}$ is estimated along with $\boldsymbol{\beta}$ and the variances of v_{it} and u_{it} in one step, using maximum likelihood.

Observation-specific estimates of efficiency can be calculated after the estimation by $\mathrm{E}\left(e^{-u_{it}} \middle| \hat{\varepsilon}_{it}\right)$. The estimates of $\boldsymbol{\beta}$ are directly interpretable as the marginal effects of the independent variables on the dependent. The same is not true for the estimates of $\boldsymbol{\xi}$. In order to get the marginal effects of the variables in $\mathbf{z}$ on efficiency one needs to calculate the derivative of $\mathrm{E}\left(e^{-u_{it}} \middle| \hat{\varepsilon}_{it}\right)$ with respect to $\mathbf{z}$. These marginal effects are usually calculated at the sample means of the data or at the specific values for every observation.

The models presented above ignore the panel nature of the data. However, when the firms differ from each other based on unobserved factors (possibly correlated with the observed data), then the pooled models run the risk of misinterpreting the unobserved heterogeneity as inefficiency. The typical solution would be to use fixed- or random-effects estimators.

In a stochastic frontier model where the group-specific constant term is assumed to be capturing the unobserved heterogeneity effect, the fixed-effects estimator would involve maximization of the likelihood function with respect to the parameter vector $\boldsymbol{\theta}$ and the set of constant terms. Although this is computationally feasible even with very large datasets (see for example Greene (2005)), typical micro-panels are short which prevents the group-specific constant terms from converging to their population counterparts. Since the estimators of the slope parameters involve the estimates of the constants, even $\boldsymbol{\theta}$ is not consistently estimated.

On the other hand, the random-effects estimator can be used to remove unobserved heterogeneity, given one is willing to assume that the heterogeneity is uncorrelated with the regressors. Greene (2005) employs the random effects estimator by simulation. However, an additional assumption is placed on the model: the group-specific constant term is assumed to have a normal density in this application, the parameters of which are to be estimated.

Here we propose treating the group-specific constant terms as missing data and then use a variation of the Monte Carlo Maximum Likelihood technique (Gelfand and Carlin, 1993) to recover the estimates of the remaining parameters. The resulting estimator relaxes both assumptions that are necessary for the random-effects model; namely, a) the independence of the unobserved heterogeneity from the regressors and b) the assumption on the distribution of the constant term. In this way, the estimator is closer to the fixed-effects estimator.

Consider the translog specification of an output distance function in inputs, outputs and time:

$$
\begin{aligned}
-\log y_{it}^{M} = {} & \alpha_i + \sum_{\ell} \beta_{\ell} \log\left[\frac{y_{it}^{\ell}}{y_{it}^{M}}\right] + \sum_{\ell}\sum_{p} \beta_{\ell p} \log\left[\frac{y_{it}^{\ell}}{y_{it}^{M}}\right] \log\left[\frac{y_{it}^{p}}{y_{it}^{M}}\right] + \sum_{\ell} \gamma_{\ell} \log x_{it}^{\ell} + \sum_{\ell}\sum_{p} \gamma_{\ell p} \log x_{it}^{\ell} \log x_{it}^{p} \\
& + \sum_{\ell}\sum_{p} \delta_{\ell p} \log\left[\frac{y_{it}^{\ell}}{y_{it}^{M}}\right] \log x_{it}^{p} + t + t^2 + \sum_{\ell} \zeta_{\ell} t \cdot \log x_{it}^{\ell} + \sum_{\ell} \eta_{\ell} t \cdot \log\left[\frac{y_{it}^{\ell}}{y_{it}^{M}}\right] + u_{it} + v_{it}
\end{aligned}
$$

where $v_{it} \sim \mathrm{N}\left(0, \sigma_v^2\right)$ and $u_{it} \sim \mathrm{N}^{+}\left(\mathbf{z}_{it}'\boldsymbol{\xi}, \sigma_u^2\right)$. For notational convenience, let y_{it} represent the dependent variable and let $\mathbf{x}_{it}$ be the vector of regressors:

$$
y_{it} = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}, \qquad \varepsilon_{it} = u_{it} + v_{it} \tag{3}
$$

Let:

a) $\sigma^2 = \sigma_v^2 + \sigma_u^2$ and $\gamma = \sigma_u^2 / \sigma^2$,

b) $d_{it} = \dfrac{\mathbf{z}_{it}'\boldsymbol{\xi}}{\sqrt{\gamma\sigma}}$ and

c) $d_{it}^{*} = (1-\gamma) d_{it} + \sqrt{\dfrac{\gamma}{1-\gamma}} \dfrac{\varepsilon_{it}}{\sigma}$

The likelihood for group i that is observed for T_i periods, conditional on the unobserved α_i is:

$$
\mathcal{L}_i = f\left(\mathbf{y}_i \middle| \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}, \alpha_i\right) = \prod_{t=1}^{T_i} \frac{1}{\sigma} \varphi\left(\frac{\varepsilon_{it} - \mathbf{z}_{it}'\boldsymbol{\xi}}{\sigma}\right) \frac{\Phi\left(d_{it}^{*}\right)}{\Phi\left(d_{it}\right)}, \qquad \varepsilon_{it} = y_{it} - \alpha_i - \mathbf{x}_{it}'\boldsymbol{\beta} \tag{4}
$$

where $\boldsymbol{\theta} = \left[\boldsymbol{\beta}' \quad \boldsymbol{\xi}' \quad \sigma \quad \gamma\right]'$ and φ and Φ are the standard normal density and distribution functions respectively.

The estimation of $\boldsymbol{\theta}$ can be based only on observed data, that is, on the density of $\mathbf{y}_i$ marginally with respect to α_i. Following Gelfand and Carlin (1993), this marginal density can be written as:

$$f\left(\mathbf{y}_i \middle| \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}\right) = f\left(\mathbf{y}_i \middle| \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}_0\right) \cdot \int_{\mathcal{A}} \frac{f\left(\mathbf{y}_i, \alpha_i \middle| \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}\right)}{f\left(\mathbf{y}_i, \alpha_i \middle| \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}_0\right)} f\left(\alpha_i \middle| \mathbf{y}_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}_0\right) \mathrm{d}\alpha_i \tag{5}$$

where $\boldsymbol{\theta}_0$ is any element of the parameter space of $\boldsymbol{\theta}$.

The joint density of $\mathbf{y}_i$ and α_i can be decomposed to the product of the known conditional density of $\mathbf{y}_i$ (from equation (4)) and the marginal density of α_i, $\mathrm{p}\left(\alpha_i \middle| \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}_0\right)$. If we assume that the density of α_i does not depend on the specific value of $\boldsymbol{\theta}$, then:

$$p\left(\alpha_i \middle| \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}\right) = p\left(\alpha_i \middle| \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}_0\right) \quad \forall \alpha_i \in \mathcal{A}$$

and (5) can be simplified to:

$$f\left(\mathbf{y}_i \middle| \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}\right) = f\left(\mathbf{y}_i \middle| \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}_0\right) \cdot \int_{\mathcal{A}} \frac{f\left(\mathbf{y}_i \middle| \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}, \alpha_i\right)}{f\left(\mathbf{y}_i \middle| \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}_0, \alpha_i\right)} f\left(\alpha_i \middle| \mathbf{y}_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}_0\right) \mathrm{d}\alpha_i \tag{6}$$

The integral in (6) is an expectation and it can be approximated by Monte Carlo methods. More specifically, the likelihood for group i can be approximated by:

$$\hat{f}\left(\mathbf{y}_i \middle| \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}\right) = f\left(\mathbf{y}_i \middle| \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}_0\right) \cdot \frac{1}{m} \sum_{j=1}^{m} \frac{f\left(\mathbf{y}_i \middle| \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}, \alpha_i^j\right)}{f\left(\mathbf{y}_i \middle| \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}_0, \alpha_i^j\right)} \tag{7}$$

where the α_i^j s are independent draws from the density of the unobserved data, given the observed and $\boldsymbol{\theta}_0$. Finally, the Monte Carlo log-likelihood function for the entire dataset is:

$$\log \hat{\mathcal{L}}\left(\boldsymbol{\theta} \middle| \mathbf{y}_i, \mathbf{x}_i, \mathbf{z}_i\right) = \sum_{i=1}^{N} \log \left[\frac{1}{m} \sum_{j=1}^{m} \frac{f\left(\mathbf{y}_i \middle| \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}, \alpha_i^j\right)}{f\left(\mathbf{y}_i \middle| \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}_0, \alpha_i^j\right)} \right] + \sum_{i=1}^{N} \log f\left(\mathbf{y}_i \middle| \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}_0\right) \tag{8}$$

and it can be maximized with respect to $\boldsymbol{\theta}$ (ignoring the second, additive term that is constant with respect to $\boldsymbol{\theta}$). Geyer (1994) proved that under loose regularity conditions, the Monte Carlo likelihood hypoconverges to the theoretical likelihood and the Monte Carlo maximum likelihood estimates converge to the maximum likelihood estimates with probability 1.

One issue that remains to be resolved is the choice of $f\left(\alpha_i \middle| \mathbf{y}_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}_0\right)$. One way to proceed would be to use Bayes' rule yielding:

$$f\left(\alpha_i \middle| \mathbf{y}_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}_0\right) \propto f\left(\mathbf{y}_i \middle|, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}_0, \alpha_i\right) \cdot \mathrm{p}\left(\alpha_i \middle|, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}_0\right)$$

Then (6) can be further simplified to:

$$f\left(\mathbf{y}_i \middle| \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}\right) = \int_{\mathcal{A}} f\left(\mathbf{y}_i \middle| \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}, \alpha_i\right) \mathrm{p}\left(\alpha_i \middle| \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}_0\right) \mathrm{d}\alpha_i \tag{9}$$

Here, we have come into full-circle and if we assume that α_i is independent of the regressors and of $\boldsymbol{\theta}_0$, the estimator becomes identical to the random-effects estimator proposed by Greene (2005). However, there is an advantage in basing simulation on (6) instead of on (9). By using Bayes' rule, as the number of time observations per group increases, the role of the prior diminishes just as it is the case in Bayesian inference. That is, after a prior $\mathrm{p}\left(\alpha_i \middle| \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}_0\right)$ is specified, when sampling from $f\left(\alpha_i \middle| \mathbf{y}_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}_0\right)$, information from the observed data is used to form the posterior density of α_i. Furthermore, even if we assume that α_i is independent of the regressors in the prior, this does not imply that it is also independent in the posterior.

3. Data

The dataset used for this paper comes from the Farm Accountancy Data Network (FADN) and covers the period 1996-2000. The dataset contains farm-level information on physical units (outputs and inputs), economic and financial data (revenues from specific products and product groups, expenses related to input use, subsidies etc.), as well as some geographical characteristics and characteristics of the farm's primary operator. FADN uses a stratified, rotating sampling scheme where farms remain in the panel on average for a period of 4-5 years.

The farms that were used for this study satisfied the following criteria:

1. they participate in the survey for at least 2 years,
2. at least 10% of the total revenues comes form cotton sales for every year that the farm is in the dataset, and
3. the revenues from sales of (i) cereals, oilseeds and proteins, (ii) vegetables, mushrooms and flowers, (iii) forage and (iv) other field crops (including cotton) comprise at least 80% of the farm's total revenues.

For the case of Greece, the last category of products is dominated by cotton and sugar beet. The other three categories contain the major products that participate in the crop rotation system. The selection of farms was made so that the resulting sample represents a rather homogeneous group in terms of their production technology. The second criterion guarantees that the farms in the used sample have integrated cotton production in the rotation system, that is, land, machinery and fixed factors in general are appropriate for the production of cotton. The third criterion selects the farms that are producing mainly arable crops. For these farms, the machinery and equipment is more likely to be homogeneous. Furthermore, farms that are producing primarily arable crops face similar decision problems on the input and output mixes and their production process could be better represented by a single distance function. The sample that resulted after applying these criteria contains 1246 farms and a total of 4371 observations.

Three outputs are used in the translog specification of the distance function: cotton production (tonnes), cereals and other output (revenues normalized by price indices). Cotton production was used as the normalizing output. The inputs are the physical units of labor and agricultural area, the cost of materials and the values of buildings and fixed equipment (buildings) and of machinery and equipment (capital).[1] Whenever only monetary values of inputs were available, the values were normalized by the corresponding price index. [2]

The specification of the mean of the truncated distribution of u_{it} includes a constant term and two sets of subsidies paid to the producers. The first set consists of the compensatory area payments paid to the producers of cereals, oilseeds, protein and energy crops. The second set contains all other subsidies. Total subsidies were divided into these two categories in order to analyze the impact of the compensatory payments on efficiency in detail.

Some additional variables were included in the specification of the inefficiency term to control for irrelevant effects on efficiency. The size of the farm in European Size Units was used to control for economies of scale in the managerial effort of the farm operators. The absolute value of investment during the accounting year (net investment plus sales of fixed factors) was included in the specification to capture the effects of possible disturbances of the production process on efficiency, in the context of the adjustment cost theory. This variable is expected to have a negative impact on a farm's efficiency score. Finally, the ratio of liabilities to total owned assets was included in the specification of the inefficiency term to control for the effects of the financial situation of the farm on efficiency. Zhengfei and Oude Lansink (2006) conclude that for the case of agriculture, liabilities should decrease inefficiency rather than increase it. This is because banks take the farm's land and buildings as collateral, thereby providing the farmers a strong incentive for being efficient.

The educational level and experience of the primary operator of the farm also belong in the specification of the inefficiency term. Two dummy variables for age of the primary operator (less than 30 and between 30 and 50 years old) are used as proxies for experience. Of course these variables are likely to capture also effects of the life cycle of the farmer. The educational level of the farmer is not available in the dataset. However, small variability for this variable is expected. According to the National Statistical Service of Greece, only 26% of the people employed in agriculture in 2000 have completed primary education.

Two models are estimated. One for farms that produce a mixture of products, including cotton but for which cotton can be treated as a regular crop participating in the rotation system. The other model uses only the farms that report at least 2/3 of the total revenues coming from sales of cotton, for every year that they participate in the FADN survey. These are the farms that specialize in cotton production. Occasionally these farms produce other output, but this happens mainly for reasons that are related to

[1] Some observations report zero values for one or more outputs or inputs. In order to avoid the biases introduced by removing these observations from the analysis, the procedure proposed by Battese (1997) was used here. In a distance function setting, the procedure implicitly imposes the restriction that the differences between the production technologies that produce all outputs (or use all inputs) and those that do not, can be captured by differences in the constant term. When an output is not produced, then the production possibilities set collapses to a smaller dimension.

[2] A detailed definition of the variables can be found in Appendix B.

the fertility of the soil. We expect the compensatory area payments to have a stronger negative impact on the efficiency scores of the farms in the second set, as the operators of these farms are more likely to view this form of support as an additional income that is disentangled from production.

Both models are estimated using the pooled stochastic frontier estimator, the linear fixed-effects and the Monte Carlo Maximum Likelihood estimator. The linear fixed-effects estimator does not provide estimates of the coefficients on the determinants of inefficiency because the inefficiency error term is absorbed by the group-specific constant. However, differences in the parameter estimates between the pooled stochastic frontier and the fixed-effects models can provide evidence for the correlation between the unobserved heterogeneity and the independent variables.

4. Results

The estimates of the distance functions are reported in Appendix A. Here we only discuss the results on the determinants of inefficiency. Table 1 presents the parameter estimates for the determinants of inefficiency from the pooled stochastic frontier models. The age of the primary operator does not have a significant effect on inefficiency for both for the mixed and the primarily cotton producers. This could be either because experience does not play an important role in determining the efficiency score of the farm, or because the effect of experience is canceled by the life-cycle effect of age. That is, older farmers may have larger experience, but on the other hand they could be less motivated to invest in new equipment, adopt innovations in the production process and in general optimize the long-run performance of the farm.

Table 1. Determinants of Inefficiency from the Pooled Stochastic Frontier Models.

	Model 1 (Mixed Producers)		**Model 2 (Cotton Producers)**	
	***nT*=3614**	***n*=962**	***nT*=1117**	***n*=284**
	Coefficient	p-value	Coefficient	p-value
age < 30	-0.0916	0.055	0.0955	0.149
30 ≤ age <50	0.0303	0.134	-0.0406	0.252
Size (ESU)	-0.0298	0.000	-0.0364	0.000
Investment	0.0023	0.382	-0.0117	0.253
Liabilities	0.4805	0.000	0.4349	0.008
Compensatory Payments	0.0544	0.000	0.1629	0.000
Other Subsidies	0.0024	0.306	-0.0113	0.002
Constant	0.3142	0.000	0.4212	0.000
σ^2	0.1177		0.0743	
γ	0.8416		0.7400	
σ_u	0.3147		0.2345	
σ_v	0.1365		0.1390	

The size of the farm reduces inefficiency and this result is in accordance with most empirical studies in agriculture. The absolute value of investment does not have a significant effect on inefficiency, while the ratio of liabilities to owned assets increases it. This could be due to the high degree of correlation between the two variable in the dataset.[3] There is an alternative explanation for the results obtained here. Instead of loans affecting the efficiency of farmers, inefficiency may induce farms to have a higher debt to assets ratio. That is, the direction of causality could be different from the one imposed by the model.

Compensatory area payments have a significant effect on efficiency in both models. As hypothesized, the effect is much stronger for the farms that are producing primarily cotton. As the area occupied by cereals, oilseeds and protein crops increases, the farms become less efficient. Furthermore, the farmers that are producing primarily cotton every year are more likely to view these area payments as an additional income, with no obligation to produce. As they rely more heavily on cotton production, they divert physical inputs and managerial effort for the production of this product, for which aid is directly related to output. This interpretation of the results is also supported by the data. By comparing the average revenues from cereals per hectare of land occupied by cereals between the two samples, one finds that for mixed producers, this quantity is €797/ha., while for the primarily cotton producers it is €601/ha.

The other subsidies appear to be reducing inefficiency for cotton producers only. As this variable contains an array of subsidies, from different support programs, this result is harder to explain. The largest portion of these subsidies comes form the set-aside premiums, which are theoretically included in the model in terms of "other output". Therefore this part of the subsidies should not affect efficiency. Another large part of these subsidies are the subsidies on inputs. The structural improvement achieved through input subsidies could result in a negative coefficient for this variable.

The marginal effects of the continuous variables on efficiency were calculated at the specific values of the variables for each observation. Table 2 presents summary statistics for these marginal effects. Again the results suggest that the compensatory area payments reduce efficiency in both groups of cotton producers, but the marginal effect is substantially larger for farms that specialize in cotton production.

Table 2. Marginal Effects of the Determinants of Inefficiency

	Model 1 (Mixed Producers)			**Model 2 (Cotton Producers)**		
	mean	min	max	mean	min	max
Size (ESU)	0.0025	0.0001	0.0033	0.0040	0.0001	0.0067
Investment	-0.0002	-0.0003	0.0000	0.0013	0.0000	0.0022
Liabilities	-0.0396	-0.0531	-0.0021	-0.0476	-0.0800	-0.0011
Comp. Payments	-0.0045	-0.0060	-0.0002	-0.0178	-0.0300	-0.0004
Other Subsidies	-0.0002	-0.0003	0.0000	0.0012	0.0000	0.0021

Table 3 presents the parameter estimates for the determinants of inefficiency from the fixed-effects stochastic frontier models. Uniform priors were assumed for the α_i's and 3,000 draws were taken from their posterior density.[4] Except for the constant term, the parameter estimates do not change

[3] When the models were estimated without the liabilities variable, investment was found to significantly increase inefficiency for mixed producers. The parameter estimate however was not significantly different from zero at the 5% level for cotton producers.

substantially from the pooled models. Many variables though become insignificant. The effects of compensatory payments become slightly larger and remain strongly significant.

Finally, when the unobserved heterogeneity is removed form the distance function, the estimates of the variance for the symmetric error terms decrease slightly. For the inefficiency error term the variance becomes larger for model 1. This comes mainly as a result in the drop of the constant term for the inefficiency model. A large portion of the fixed inefficiency that was captured by the constant term in the pooled models is removed along with the unobserved heterogeneity. The shape of the distribution of u changes to accommodate observations that correspond to highly inefficient units.

Table 3. Determinants of Inefficiency from the Monte Carlo ML Models.

	Model 1 (Mixed Producers)		**Model 2 (Cotton Producers)**	
	***nT*=3614**	***n*=962**	***nT*=1117**	***n*=284**
	Coefficient	p-value	Coefficient	p-value
age < 30	-0.2154	0.207	0.1697	0.245
30 ≤ age <50	0.0737	0.165	-0.0548	0.496
Size (ESU)	-0.0359	0.000	-0.0348	0.005
Investment	0.0089	0.108	-0.0005	0.980
Liabilities	0.7063	0.000	0.3961	0.262
Compensatory Payments	0.0669	0.000	0.1800	0.000
Other Subsidies	0.0003	0.929	-0.0070	0.339
Constant	-0.2795	0.027	0.0832	0.463
σ^2	0.1449		0.0637	
γ	0.9224		0.8406	
σ_u	0.3656		0.2314	
σ_v	0.1061		0.1008	

5. Concluding remarks

This paper analyzed the effects of compensatory area payments on the technical efficiency scores of cotton producers in Greece. Using a distance function representation of the production technology and accounting for unobserved heterogeneity in its estimation, we were able to better identify these effects. As most cotton producers also produce many of the products for which support is partially decoupled from production, we found that compensatory area payments significantly reduce efficiency. Decoupled aid can be viewed by farmers as an alternative source of income additional to the sales of cotton. Although land is allocated for the production of cereals, no considerable effort is made to exploit the full potential of the production technology. On the contrary, resources are reallocated for production of larger volume of the cash crop. The negative effects of the compensatory area payments on efficiency were found to be much stronger for producers that rely heavily on cotton production as a

[4] The variance-covariance matrix was calculated using the BHHH estimator of the information matrix: $\hat{\mathbf{I}}(\hat{\boldsymbol{\theta}}) = \sum_i \mathbf{s}_i(\hat{\boldsymbol{\theta}})\mathbf{s}_i'(\hat{\boldsymbol{\theta}})$, where $\mathbf{s}_i(\hat{\boldsymbol{\theta}})$ is the score of the simulated likelihood function for observation i. The Monte Carlo standard error was ignored during the calculation.

source of income. Finally, we find that ignoring the presence of unobserved heterogeneity will overstate the levels of inefficiency.

References

Aigner, D., Lovell, C. A. K., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production functions. *Journal of Econometrics*, 6:21–37.

Battese, G.E., (1997). A note on the estimation of Cobb–Douglas production functions when some explanatory variables have zero values. *Journal of Agricultural Economics* 48, 250–252.

Battese, G. E. and Coelli, T. J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*, 20:325–332.

Battese, G. E. and Corra, G. S. (1977). Estimation of a production frontier model: With application to the pastoral zone of eastern Australia. *Australian Journal of Agricultural Economics*, 21:169–179.

Bezlepkina, I., Oude Lansink, A. and Oskam, A. (2005). Effects of subsidies in Russian dairy framing. *Agricultural Economics*, 33:277–288.

Coelli, T. and Perelman, S. (1996). Efficiency measurement, multiple-output technologies and distance functions: with application to European railways. CREPP Discussion Paper No. 96/05, Univeristè de Liège, Liège.

Gelfand, A. E. and Carlin, B. P. (1993). Maximum-likelihood estimation for constrained or missing-data models. *The Canadian Journal of Statistics*, 21:303–311.

Geyer, C. J. (1994). On the convergence of Monte Carlo maximum likelihood calculations. *Journal of the Royal Statistical Society*, Series B 56:261–274.

Greene, W. H. (1990). A gamma distributed stochastic frontier model. *Journal of Econometrics*, 46:141–164.

Greene, W. H. (2005). Fixed and random effects in stochastic frontier models. *Journal of Productivity Analysis*, 23:7–32.

Guyomard, H., Baudry M., and Carpentier, A. (1996). Estimating crop supply response in the presence of farm programmes: application to the CAP. *European Review of Agricultural Economics*. 23:401–420.

Hennessy, D.A. (1998). The production effects of agricultural income support polices under uncertainty. *American Journal of Agricultural Economics*, 80:46–57.

Karagiannis, G and Pantzios, C. (2002). To comply or not to comply with policy regulations- the case of Greek cotton growers: a note. *Journal of Agricultural Economics*, 53:345–351.

Kleinhanß, W., Murillo C., San Juan. C and Sperlich S. (2007). Efficiency, subsidies, and environmental adaptation of animal farming under CAP. *Agricultural Economics*, 36:49–65.

Kumbhakar, S. C., Ghosh, S. and McGuckin, J. T. (1991). A generalized production frontier approach for estimating determinants of inefficiency in U.S. dairy farms. *Journal of Business and Economic Statistics*, 9:279–286.

Meeusen, W. and van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, 18:435–444.

Moro,D., and P. Sckokai. (1999). Modelling the CAP arable crop regime in Italy: Degree of decoupling and impact of Agenda 2000. *Cahiers d'Economie et Sociologie Rurales* 53:49–73.

Oude Lansink, A. and Peerlings, J. (1996). Modelling the new EU cereals regime in the Netherlands. *European Review of Agricultural Economics*, 23: 161-178.

Reifschneider, D. and Stevenson, R. (1991). Systematic departures from the frontier: a framework for the analysis of firm inefficiency. *International Economic Review*, 32:715–723.

Sckokai, P. and Moro, D. (2006) Modeling the reforms of the Common Agricultural Policy for arable crops under uncertainty. *American Journal of Agricultural Economics*. 88:43–56.

Stefanou, S.E. and Saxena, S. (1988). Education, experience and allocative efficiency: a dual approach. *American Journal of Agricultural Economics* 70:338–345.

Stevenson, R. E. (1980). Likelihood functions for generalized stochastic frontier estimation. *Journal of Econometrics*, 13:57–66.

Zhengfei, G. and Oude Lansink, A. (2006). The source of productivity growth in Dutch agriculture: a perspective from finance. *American Journal of Agricultural Economics*, 88:644–656.

Appendix A

Table A1. Parameter Estimates for the Distance Function

	Model 1 (Mixed Producers)			Model 2 (Cotton Producers)		
	Pooled Stochastic Frontier	**Linear Fixed Effects**	**Monte Carlo ML**	**Pooled Stochastic Frontier**	**Linear Fixed Effects**	**Monte Carlo ML**
log_cer	0.2313	-0.2065	-0.1812	0.2770	0.2054	0.2937
log_oth	-0.1302	-0.0556	-0.0061	0.2125	0.3215	0.2420
log_K	-0.0972	0.1228	-0.0193	0.1286	-0.2107	-0.1474
log_B	0.0506	0.3247	0.4066	0.1382	0.3952	0.3295
log_L	-1.8136	-1.2530	-1.3029	-0.5203	-0.1282	-0.4609
log_A	2.4346	1.3023	1.4965	1.7816	1.1945	0.9651
log_M	-1.7869	-1.4740	-1.4245	-2.5205	-1.6322	-1.3380
log_KK	-0.0042	-0.0039	-0.0064	-0.0029	0.0054	0.0070
log_KB	-0.0018	-0.0039	-0.0051	0.0009	-0.0029	-0.0059
log_KL	-0.0308	-0.0102	-0.0073	-0.0171	-0.0130	-0.0220
log_KA	-0.0413	0.0117	-0.0074	-0.0010	-0.0120	-0.0023
log_KM	0.0527	-0.0031	0.0179	0.0008	0.0230	0.0166
log_BB	0.0005	-0.0091	-0.0122	-0.0061	-0.0234	-0.0201
log_BL	-0.0058	-0.0077	-0.0082	0.0046	0.0100	0.0074
log_BA	0.0057	0.0105	0.0159	0.0136	0.0254	0.0271
log_BM	-0.0002	-0.0115	-0.0150	-0.0114	-0.0183	-0.0123
log_LL	0.0550	0.0482	0.0467	-0.0259	-0.0032	0.0320
log_LA	-0.2556	-0.1582	-0.1369	-0.0719	-0.0626	-0.0786
log_LM	0.1847	0.0934	0.0956	0.1175	0.0391	0.0310
log_AA	0.2147	0.0852	0.1454	0.1906	0.0469	0.0153
log_AM	-0.2096	-0.1558	-0.2059	-0.3164	-0.1886	-0.1437
log_MM	0.0220	0.0624	0.0558	0.1266	0.0860	0.0708
log_cc	0.0372	0.0456	0.0443	0.0104	0.0303	0.0161
log_co	-0.0049	-0.0070	-0.0057	-0.0005	0.0010	0.0019
log_oo	0.0494	0.0394	0.0351	0.0055	0.0081	0.0101
log_Kc	-0.0009	0.0009	0.0026	-0.0044	-0.0034	-0.0036
log_Bc	-0.0028	-0.0025	-0.0031	-0.0005	-0.0031	-0.0016
log_Lc	0.0121	0.0135	0.0119	-0.0075	-0.0294	-0.0061
log_Ac	0.0698	0.0229	0.0214	0.0345	0.0657	0.0639
log_Mc	-0.0648	-0.0149	-0.0163	-0.0275	-0.0214	-0.0409
log_Ko	0.0017	0.0091	0.0078	-0.0074	-0.0167	-0.0141
log_Bo	-0.0010	-0.0009	-0.0006	0.0001	0.0018	0.0000
log_Lo	-0.0151	-0.0200	-0.0211	0.0073	0.0090	0.0045
log_Ao	0.0292	0.0340	0.0315	0.0373	0.0540	0.0449
log_Mo	-0.0169	-0.0192	-0.0175	-0.0291	-0.0410	-0.0292
t	-0.0819	0.0386	0.0216	-0.3304	-0.2758	-0.2909
t2	0.0056	-0.0010	0.0028	0.0198	0.0209	0.0218
tlog_cer	0.0037	0.0030	0.0026	0.0044	0.0045	0.0027

	Model 1 (Mixed Producers)			Model 2 (Cotton Producers)		
	Pooled Stochastic Frontier	Linear Fixed Effects	Monte Carlo ML	Pooled Stochastic Frontier	Linear Fixed Effects	Monte Carlo ML
tlog_oth	0.0054	0.0048	0.0044	0.0018	0.0001	0.0021
tlog_K	-0.0046	-0.0053	-0.0040	0.0086	0.0117	0.0137
tlog_B	-0.0009	-0.0023	-0.0028	-0.0010	-0.0006	-0.0007
tlog_L	0.0380	0.0312	0.0312	0.0164	-0.0011	-0.0005
tlog_A	0.0086	0.0226	0.0186	-0.0167	-0.0108	-0.0129
tlog_M	-0.0310	-0.0361	-0.0361	-0.0013	0.0006	0.0000
d_cer	-0.1756	-0.2109	-0.1906	0.0549	-0.0016	0.0797
d_oth	-0.5103	-0.3123	-0.2061	0.1371	0.0459	0.0822
d_bldg	-0.0620	0.4745	0.6350	0.4350	1.2231	1.0124
constant	11.5479			8.9352		

Appendix B

Table B1. Description of the variables used in the model.

cotton	Cotton production in tonnes.
cereals	Revenues from sales of cereals (excluding rice) divided by the annual price index for cereals.
other	Revenues from sales of other products divided by the annual price index for agricultural products.
log_K	Natural logarithm of the value of machinery and equipment (machines, tractors, cars, etc.) divided by the annual price index for agricultural machinery and installations.
log_B	Natural logarithm of the value of buildings and fixed equipment belonging to the holder divided by the annual price index for farm buildings.
log_L	Natural logarithm of the time worked in hours.
log_A	Natural logarithm of the total utilized agricultural area in hectares.
log_M	Natural logarithm of the value of the materials used (fertilizers, energy, etc.) divided by the annual price index for materials in agriculture.
d_cereals	dummy variable: 1 if the farm does not produce cereals.
d_other	dummy variable: 1 if the farm does not produce "other" output.
d_bldg	dummy variable: 1 if the farm reports zero value of owned buildings.
Size	Economic size of the farm in European Size Units
Investment	Absolute value of the value of total investment during the accounting year (including investment plus sales of capital and buildings)
Liabilities	Ratio of total liabilities to total (owned) assets
Compensatory Payments	Amounts paid to producers of cereals, oilseeds and protein, and energy crops in 1,000's Euros.
Other Subsidies	Total subsidies related to production paid to farmers, excluding compensatory payments, in 1,000's Euros. Includes set-aside premiums, subsidies on inputs, LFA and environmental subsidies.